

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

**Speech Controls For Use With a Speech System**

Inventors:

Stephen R. Falcon

Clement Yip

Dan Banay

David M. Miller

ATTORNEY'S DOCKET NO. MS1-1007US

## **TECHNICAL FIELD**

The systems and methods described herein relate to speech systems. More particularly, the described invention relates to speech controls for use by speech-enabled applications with a speech system.

## **BACKGROUND**

Speech systems have been incorporated into many useful applications so that users may utilize the applications without having to manually operate an input device, such as a mouse or a keyboard. Personal computer systems (desktop, laptop, handheld, *etc.*) and automobile systems are only two examples of systems, or platforms, that may include integrated speech recognition functions.

A single platform may have several applications executing at a given time. For example, in an automobile computer system that utilizes speech recognition software, there may be speech-enabled applications for radio operation, navigational tools, climate controls, mail, *etc.* Personal computers may include word processors, spreadsheets, databases and/or other programs that utilize speech recognition. Each speech-enabled application has a grammar associated with it that is a set of commands that the application is attempting to detect at any one time.

Different applications may have different grammars. For instance, a word processing speech-enabled application may use a grammar that enables it to detect the command "print." However, an automobile speech-enabled application that controls a car radio would not have such a command. On the other hand, the car radio application may have a grammar that enables the speech system to recognize

1 the command "FM" to set the radio to the FM band. The word processor would  
2 not waste overhead by including an "FM" command in its relevant grammar.

3 As the number of speech-enabled applications and grammars has increased,  
4 it has become increasingly problematic to run multiple speech-enabled  
5 applications on a single platform. Although each speech-enabled application may  
6 have its own unique grammar, certain commands may be used in more than one  
7 grammar, *e.g.*, "stop." When a speech system receives such a command, it must  
8 be able to determine which application the speaker directed the command to and  
9 which application should respond to the user.

10 Similarly, multiple speech-enabled applications may attempt to deliver  
11 speech feedback simultaneously. This can result in a garbled communication that  
12 a user cannot understand. Such a result renders one or more of the applications  
13 useless. Also, if speech feedback from one speech-enabled application interrupts  
14 speech feedback from another similar application, the feedback from one or both  
15 applications may not be understandable to a user.

16 For example, suppose a first application asks a question of the user and  
17 awaits a response. But before the user responds to the first application, a second  
18 application asks the user a question. Which application will accept the user's first  
19 answer? Will one of the applications accept an answer intended for the other  
20 application? Will either application be able to function properly with the  
21 response(s) it receives? With no control over specific interactions between the  
22 system and the user, there is no certain answer to any of these questions.

23 One method that has been devised to handle this problem is to create a  
24 'token' that indicates which application has the right to execute at any given time.  
25

1 When an application is ready to execute it requests a token. When the application  
2 receives the token, the application may execute.

3 One of several drawbacks of such a system is that applications may crash or  
4 hang. If an application that currently holds the token crashes, then the system may  
5 not recover unless the system is prepared for application crashes. If the  
6 application hangs, then the system may never be able to regain control. Therefore,  
7 a token system is an inadequate solution to the problems encountered when  
8 attempting to execute multiple speech-enabled applications.

9 Another problem that is encountered by speech-enabled applications is that  
10 when a command is given to an application that is not currently running, the  
11 command simply falls on deaf ears, so to speak, and there is no response to the  
12 command. Therefore, a user must first manually or vocally launch the application,  
13 then speak the desired command for the application. This means that a user must  
14 always be aware of which applications are running and which are not, so that the  
15 user knows whether she must launch an application before issuing certain  
16 commands. For example, if an automobile driver wants to play "song\_A.mp3" on  
17 a car radio, the driver must first issue a command or manually launch an MP3  
18 player, then command the player to play "song\_A." It would be desirable to  
19 minimize the actions required to launch an application and subsequently issue a  
20 command.

## SUMMARY

Systems and methods are described for providing speech controls to speech-enabled applications to use with a speech system. Providing the speech controls to speech-enabled application developers prevents a developer from having to write similar code each time the developer wants a speech-enabled application to perform a particular function. The speech controls described herein are a reusable set of controls for common interaction scenarios in speech-enabled applications. A speech-enabled application may include from one to all of the described speech controls.

Use of the speech controls across various speech-enabled applications provides consistency for users. For example, a user has a speech recognition computer in his car. If the user buys a new car that has a speech recognition computer, the user is assured that many of the functions will operate like they did in his old car.

The speech controls also provides consistency between applications, such as in automobiles. As a result, a car owner finds it easier to use various speech-enabled applications in her car because many of the interactions between the car owner and the applications are standardized.

In one implementation, a question control is described that provides a specific format for asking questions requesting user input. The question control gives an application developer an easy way to create various system-initiated interactions, or dialogues.

The question control allows flexible programming so that a variety of question scenarios can be implemented. For example, the question control may be used to ask a driver a simple question that may be answered "yes" or "no", or a

1 more complex question such as "fast or scenic route" and receive "fast" or  
2 "scenic" as appropriate answers.

3 In another implementation, an announcer control is described. The  
4 announcer control provides a developer an easy way to deliver verbal feedback to  
5 users, including short notices and long passages of text-to-speech. The announcer  
6 control implements a simple mechanism for playing canned speech or TTS text,  
7 and for giving a user standardized control of such playback. Use of the announcer  
8 control significantly decreases the effort required by application developers to  
9 build a rich application user interface.

10 A command control is described in one implementation. The command  
11 control is designed to easily attach command-and-control grammar to an  
12 application. The command control is used for user-initiated speech. The  
13 command control performs at least two functions. First, the command control  
14 provides a way for an application to specify what grammar(s) the application is  
15 interested in listening to. Second, the command control communicates back to the  
16 application that a recognition has occurred.

17 In another implementation, a word trainer control is described. The word  
18 trainer control provides an efficient way to implement a speech-oriented word-  
19 training interaction with a user, in support of tasks that involve voice tags, such as  
20 speed-dial entries or radio station names. The entire word training process is  
21 implemented with a combination of the word trainer control and other GUI  
22 (graphical user interface) or SUI (speech user interface) controls. The word  
23 trainer primarily focuses on the audio recording portions of the process.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

A more complete understanding of exemplary methods and arrangements of the present invention may be had by reference to the following detailed description when taken in conjunction with the accompanying drawings wherein:

Fig. 1 is a block diagram of a computer system conforming to the invention described herein.

Fig. 2a is a diagram of an exemplary interaction.

Fig. 2b is a diagram of an exemplary interaction.

Fig. 2c is a diagram of an exemplary interaction.

Fig. 3 is a flow diagram depicting a methodological implementation of interaction processing.

Fig. 4 is a flow diagram depicting a methodological implementation of interaction interruption.

Fig. 5 is a flow diagram depicting a methodological implementation of interaction chaining.

Fig. 6 is a flow diagram depicting a methodological implementation of chained interaction interruption.

Fig. 7 is a flow diagram depicting a methodological implementation of grace period interruption.

Fig. 8a is a diagram of an exemplary master grammar table.

Fig. 8b is a diagram of an exemplary grammar table and its components.

Fig. 9 is a diagram of an exemplary computing environment within which the present invention may be implemented.

Fig. 10 is a flow diagram of a question control process.

Fig. 11 is a flow diagram of an announcer control process.

1 Fig. 12a is a block diagram of a command manager control.

2 Fig. 12b is a representation of a command manager object interface.

3 Fig. 13 is a representation of a speech server interface.

## 4 5 **DETAILED DESCRIPTION**

6 This invention concerns a speech system that is able to manage interactions  
7 from multiple speech-enabled applications to facilitate meaningful dialogue  
8 between a user and the speech system. This invention speech system may be  
9 applied to a continuous speech system as well as a discrete speech system.

10 Furthermore, the invention may be described herein as an automobile  
11 speech system or systems. However, the invention may also be implemented in  
12 non-automobile environments. Reference may be made to one or more of such  
13 environments. Those skilled in the art will recognize the multitude of  
14 environments in which the present invention may be implemented.

### 15 General Terms

16 Following is a brief description of some of the terms used herein. Some of  
17 the terms are terms of art, while others are novel and unique to the described  
18 invention. Describing the terms initially will provide proper context for the  
19 discussion of the invention, although the descriptions are not meant to limit the  
20 scope of the terms in the event that one or more of the descriptions conflict with  
21 how the terms are used in describing the invention.

### 22 Grammars

23 As previously stated, each speech-enabled application likely has its own  
24 specific grammar that a speech system must recognize. There are a variety of  
25 different things that applications will want to do with their grammars, such as



1 constructing new grammars, using static grammars, enable/disable rules or entire  
2 grammars, persist grammars, make the grammars continually available, etc. The  
3 speech system described herein exposes methods to accomplish these things and  
4 more.

5 Different grammars can have different attributes. A static grammar is one  
6 that will not change after being loaded and committed. A dynamic grammar, to  
7 the contrary, is a grammar that may change after a commit. Whether a grammar is  
8 static or dynamic must be known when the grammar is created or registered with  
9 the speech system. Rules may also be static or dynamic. A static rule cannot be  
10 changed after it is committed, while a dynamic rule may be changed after it is  
11 committed. A static rule can include a dynamic rule as a part of the static rule.

12 A grammar may, at any time, be an enabled grammar or a disabled  
13 grammar. A disabled grammar is still within the speech system, but is not being  
14 listened for by the system. An enabled grammar may also be called an active  
15 grammar; a disabled grammar may also be referred to as an inactive grammar.

16 Reference is made herein to transient and persistent grammars. A transient  
17 grammar is a grammar that is only active while its corresponding application is  
18 executing. When the application halts execution, i.e., shuts down, the grammar is  
19 removed from the speech system. A persistent grammar is always present in the  
20 speech system, whether the application to which the grammar belongs is present in  
21 the system. If an utterance is heard that belongs to a persistent grammar and the  
22 application is not running to handle it, the speech system launches the application.

23 Furthermore, reference is made herein to global and yielding grammars. A  
24 global grammar contains terms that the speech system is always listening for.  
25 Global grammars are used sparingly to avoid confusion between applications. An

1 example of a global grammar is a "call 9-1-1" command. A yielding grammar is  
2 active unless another grammar takes focus. The reason that another grammar  
3 would take focus is that a conversation unrelated to the grammar becomes active  
4 and yielding grammars outside the conversation are disabled.

### 5 Interaction

6 The term "interaction" is used herein to refer to a complete exchange  
7 between a speech-enabled application and a user. An interaction is a context of  
8 communication that unitizes one or more elements of a dialogue exchange. For  
9 example, an application developer may want to program a speech-enabled  
10 application to alert a user with a tone, ask the user a question, and await a response  
11 from the user. The developer would likely want these three events to occur  
12 sequentially, without interruption from another application in order for the  
13 sequence to make sense to the user. In other words, the developer would not want  
14 the alert tone sounded and the question asked only to be interrupted at that point  
15 with a communication from another application. The user may then not know how  
16 or when to respond to the question. Therefore, with the present invention, the  
17 developer may include the three actions in one interaction that is submitted to a  
18 speech sequential execution. Only in special circumstances will an interaction be  
19 interrupted. Interactions will be discussed in greater detail below.

### 20 Conversation

21 A series of related interactions may be referred to herein as a  
22 "conversation." A conversation is intended to execute with minimal interruptions.

### 23 Computer-Executable Instructions/Modules

24 The invention is illustrated in the drawings as being implemented in a  
25 suitable computing environment. Although not required, the invention is

1 described in the general context of computer-executable instructions, such as  
2 program modules, to be executed by a computing device, such as a personal  
3 computer or a hand-held computer or electronic device. Generally, program  
4 modules include routines, programs, objects, components, data structures, etc. that  
5 perform particular tasks or implement particular abstract data types. Moreover,  
6 those skilled in the art will appreciate that the invention may be practiced with  
7 other computer system configurations, including multi-processor systems,  
8 microprocessor-based or programmable consumer electronics, network PCs,  
9 minicomputers, mainframe computers, and the like. The invention may also be  
10 practiced in distributed computing environments where tasks are performed by  
11 remote processing devices that are linked through a communications network. In  
12 a distributed computing environment, program modules may be located in both  
13 local and remote memory storage devices.

#### 14 Exemplary Speech System

15 Fig. 1 is a block diagram of a computer system 100 that includes a speech  
16 system 102 and memory 104. The computer system 100 also includes a processor  
17 106 for executing computer instructions, a display 108, an input/output (I/O)  
18 module 110, a speaker 112 for speech output, a microphone 114 for speech input,  
19 and miscellaneous hardware 116 typically required in a computer system 100. The  
20 computer system 100 may be designed for use in an automobile or in a non-  
21 automobile environment, such as in a desktop computer, a handheld computer, an  
22 appliance, etc.

23 The speech system 100 includes a speech engine 118 having a text-to-  
24 speech (TTS) converter 120 and a speech recognizer (SR) 122. The TTS  
25 converter 120 and the speech recognizer 122 are components typically found in

1 speech systems. The speech recognizer 122 is configured to receive speech input  
2 from the microphone 114 and the TTS converter 120 is configured to receive  
3 electronic data and convert the data into recognizable speech that is output by the  
4 speaker 112.

5 The speech system 102 also includes a speech server 124 that  
6 communicates with the speech engine 118 by way of a speech application  
7 programming interface (SAPI) 126. Since the speech engine 118 is separate from  
8 the speech server 124, the speech server 124 can operate with any number of  
9 vendor-specific speech engines via the speech API 126. However, such a specific  
10 configuration is not required.

11 The SAPI 126 includes a vocabulary 164 that is the entire set of speech  
12 commands recognizable by the speech system 102. It is noted that speech engine  
13 118 may include the vocabulary 164 or a copy of the vocabulary 164 that is  
14 contained in the SAPI 126. However, the present discussion assumes the  
15 vocabulary 164 is included in the SAPI 126.

16 Several applications may be stored in the memory 104, including  
17 application\_1 130, application\_2 132 and application\_n 134. Depending on the  
18 components that make up the computer system 100, virtually any practical number  
19 of applications may be stored in the memory 104 for execution on the speech  
20 server 124. Each application 130 -134 is shown including at least one control:  
21 Application\_1 130 includes a question control 154; application\_2 includes an  
22 announcer control 156; and application\_n includes a command control 156 and a  
23 word trainer control 158.

24 Each control 154 – 158 uses a specific grammar: the question control 154  
25 uses grammar\_1 136; the announcer control 156 uses grammar\_2 138; the

1 command control 156 uses grammar\_3 152; and the word trainer control 158 uses  
2 grammar\_4 140.

3 The controls 154 - 158 are designed to provide application developers a  
4 robust, reliable set of user-interface tools with which to build applications. The  
5 controls 154 - 158 are code modules that perform recurring functions desired by  
6 application developers. The controls 154 - 158 decrease the programming effort  
7 required by an original equipment manufacturer or an independent vendor to  
8 create a rich application user interface.

9 The question control 154 gives an application developer an easy way to  
10 create various modal, system-initiated interactions, or dialogues. The announcer  
11 control 155 provides a developer a simple way to deliver verbal feedback to users,  
12 including short notices and long passages of text-to-speech. The command control  
13 156 provides a way for applications to specify what grammar it is interested in  
14 listening to, and communicates to the applications if and when a recognition  
15 occurs. The word trainer control 158 provides an easy way to implement a  
16 speech-oriented work-training interaction with a user. These controls will be  
17 discussed in greater detail below.

18 It is noted that the speech server 126 and the applications 130 - 134 are  
19 separate processes. In most modern operating systems, each process is isolated  
20 and protected from other processes. This is to prevent one application from  
21 causing another application that is running to crash. A drawback with utilizing  
22 separate processes is that it makes sharing data between two processes difficult,  
23 which is what the speech server 126 needs to do in this case. Therefore, data must  
24 be marshaled between the applications 130 - 134 and the speech server 126.  
25

1 There are various ways to marshal data across process boundaries and any  
2 of those ways may be used with the present invention. A common way to marshal  
3 data is with the use of a proxy and a stub object. A proxy resides in the  
4 application process space. As far as the proxy is concerned, the stub object is the  
5 remote object it calls. When an application calls some method on a proxy object,  
6 it does so internally, which is necessary to package data passed by the application.  
7 into the speech server process space, the stub object receives the data and calls a  
8 target object in the speech server. However, it is noted that any method known in  
9 the art to marshal data between processes may be used.

10 The speech server 124 also includes an interaction manager 160, a master  
11 grammar table 164 and a speech server interface 148. The master grammar table  
12 162 contains one or more grammars that are registered with the speech server 124  
13 by one or more applications. The master grammar table 162 and the registration  
14 of grammars will be discussed in greater detail below, with reference to Fig. 3.

15 The interaction manager 160 maintains an interaction list 168 of one or  
16 more interactions (interaction\_1 170, interaction\_2 172, interaction\_3 174,  
17 interaction\_n 176) from one or more applications in a particular order for  
18 processing by the speech server 124. As previously discussed, an interaction is a  
19 logical context used by an application to communicate with a user. At any given  
20 time, there can be, at most, one active interaction between the user and an  
21 application. The interaction manager 160 processes the interactions 170 - 176 in  
22 order. Interactions can be inserted at the front of the interaction list 168, i.e.,  
23 before interaction\_1 170, or at the end of the interaction list 168, i.e.,  
24 interaction\_n. If an interaction is inserted at the front of the interaction list 168,  
25 the processing of interaction\_1 170 will be interrupted. In one implementation,

1 the interrupting interaction will only interrupt a current interaction if the  
2 interrupting interaction is configured to take precedence over a currently executing  
3 interaction.

4 The interaction manager 160 is also configured to notify the applications  
5 170 -176 of the following transitions so that the applications 170 - 176 may  
6 modify the state or content of an interaction as it is processed in the interaction list  
7 168: interaction activated, interaction interrupted, interaction self-destructed,  
8 interaction re-activated, and interaction completed. As a result, the applications  
9 170 - 176 can be aware of the state of the speech system 102 at all times.

10 As previously noted, an interaction contains one or more elements that  
11 represent a “turn” of communication. A turn is a single action taken by either the  
12 system of the user during an interaction. For example, the system may announce  
13 “Fast or scenic route?” during a turn. In response, the user may answer “Fast,”  
14 which is the user’s turn.

### 15 Exemplary Interactions

16 Fig. 2 illustrates some examples of interactions. Fig. 2a depicts exemplary  
17 interaction\_A 200. Interaction\_A 200, when executed, will sound a tone, ask a  
18 question and await a response from a user. Interaction\_A 200 includes three  
19 elements that each represent a turn of communication; the first turn is the tone, the  
20 second turn is the question, and the third turn is the waiting. The first element is  
21 an EC (earcon) 210, which causes an audio file to be played. In this example, the  
22 EC 210 sounds a tone to alert a user that the speech system 102 is about to ask the  
23 user a question. The second element is a TTS (text-to-speech) 212 element that  
24 plays a text file (i.e., speaks), which in this example, asks the user a question. The  
25 last element is an SR (speech recognition) 214 element that listens for a term

1 included in the vocabulary 164, Fig. 1. Processing exemplary interaction\_A 200  
2 creates the desired result from the speech system 102.

3 Fig. 2b depicts exemplary interaction\_B 220 that also includes three  
4 elements: an EC 222, a TTS 224 and a WT (word trainer) 226 element.  
5 Processing interaction\_B 226 results in the speech system sounding a tone, asking  
6 the user to state a command, and assigns the response stated by the user to a  
7 command.

8 Fig. 2c depicts exemplary interaction\_C 230 that includes two elements: a  
9 TTS 232 and an EC 234. Processing interaction\_C 230 results in the speech  
10 system 102 playing a text file followed by the playing of an audio file.

11 There is another type of element (not shown) that may be inserted into an  
12 interaction to cause a delay, or time out, before the system processes subsequent  
13 elements. This type of element is referred to as a NULL element. A NULL  
14 element would be inserted into an interaction to allow additional time for the  
15 interaction to be processed.

16 Referring now back to Fig. 1, the interaction manager 160 provides for the  
17 ordering of interactions, including the elements (EC, TTS, WT, NULL, SR)  
18 discussed above. This prevents more than one application from addressing the  
19 user simultaneously. The interaction manager 160 processes the interactions 170 -  
20 176 in the interaction list 168 in the order in which the interactions are submitted  
21 to the interaction manager 160 (i.e., on a first-in-first-out basis). An exception to  
22 this is that an application is provided the ability to submit an interaction directly to  
23 the beginning of the interaction list 168 in situations where the application  
24 considers the interaction a high priority.

## 25 Interaction Management: Methodological Implementation



1 Fig. 3 is a flow diagram depicting a way in which the interaction manager  
2 168 functions to manage the interactions 170 - 176 in the interaction list 168. In  
3 the discussion of Fig. 3 and the following figures, continuing reference will be  
4 made to the features and reference numerals contained in Fig. 1.

5 At block 300, interaction\_A 170 is active, while interaction\_B 172 and  
6 interaction\_C 174 wait in the interaction list 168 to be processed. At block 302,  
7 interaction\_n 176 is added to the end of the interaction list 168. Interaction\_A 170  
8 continues processing at block 304 ("No" branch, block 306) until it concludes.  
9 Then, interaction\_B 172 becomes active, i.e., begins processing at block 308  
10 ("Yes" branch, block 306).

11 Interruption occurs when an application places an interaction at the  
12 beginning of the interaction list 168 without regard to an interaction already active  
13 there. When an interruption occurs, the active interaction is deactivated, and the  
14 interrupting interaction is activated.

#### 15 Interaction Interruption: Methodological Implementation

16 Fig. 4 is a flow diagram depicting an interaction interruption. On the left  
17 side of the figure, a current state of the interaction list 168 is shown corresponding  
18 to the blocks contained in the flow diagram. At block 400, interaction\_A 170 is  
19 active while interaction\_B 172 and interaction\_C 174 are inactive and waiting in  
20 the interaction list 168 to be processed. While interaction\_A 170 is executing,  
21 interaction\_n 176 is submitted by one of the speech-enabled applications 130 - 134  
22 (block 402). The submitting application wants interaction\_n 176 to be processed  
23 immediately without regard to other interactions in the interaction list 168, so an  
24 interruption flag is set in interaction\_n 176 that tells the interaction manager 160  
25 to process interaction\_n 176 right away.

1 Interaction\_n 176 is then processed at block 406 ("No" branch, block 404)  
2 until it has completed, i.e., actions related to any and all elements contained in  
3 interaction\_n 176 have been performed. Only when interaction\_n 176 has  
4 completed processing ("Yes" branch, block 404), does interaction\_A 170 have the  
5 capability to process again.

6 However, interactions submitted to the interaction list 168 have a self-  
7 destruct option that, when used, terminates the interaction in the event that the  
8 interaction is interrupted. In some cases, an interaction may need to self-destruct  
9 due to internal failure. In such cases, the situation is treated the same as a normal  
10 self-destruction.

11 At block 408, it is determined whether interaction\_A 170 has set a self-  
12 destruct flag that indicates the interaction should self-destruct upon interruption.  
13 If the self-destruct flag (not shown) is set ("Yes" branch, block 408), interaction\_A  
14 170 terminates (block 410). If the self-destruction flag is not set ("No" branch,  
15 block 408), then interaction\_A 170 finishes processing at block 412.

16 Interactions do not have an inherent "priority." The applications only have  
17 an absolute ability to place an interaction at the front of the interaction list 168.  
18 Such a placement results in interruption of a current interaction being processed.

19 In another implementation, not shown, an interrupting interaction will not  
20 be processed until a current interaction has concluded if the current interaction will  
21 conclude in a pre-determined period of time. If the current interaction will take a  
22 longer amount of time than the pre-determined time to conclude, it is interrupted  
23 as described above.

24 For example, suppose that an interrupting interaction will only interrupt a  
25 current interaction if the current interaction will not complete within three

1 seconds. If a driver is just completing a long interaction that has taken thirty  
2 seconds to process but will conclude in two seconds, it may be desirable to let the  
3 interaction finish before interrupting with, say, an engine overheating  
4 announcement. If the current interaction is not self-destructing, the driver may  
5 have to endure another thirty-two seconds of interaction that he doesn't want to  
6 hear if the current interaction is repeated after the overheating announcement  
7 concludes. This would become even more irritable if another engine overheating  
8 announcement interrupted the current interaction again and the current interaction  
9 repeated again.

#### 10 Interaction Chaining: Methodological Interaction

11 Interactions may also be "chained" together by speech-enabled applications  
12 using the speech server 124. An application may want a certain interaction to  
13 establish a chain of interactions that constitutes a conversation. When this is the  
14 case, when an interaction concludes processing, the speech server 124 will wait a  
15 pre-determined grace period, or time out, before processing the next interaction in  
16 the interaction list 168. During the grace period, the application may submit a  
17 subsequent interaction.

18 An example of when interaction chaining may be used is when an  
19 automobile navigation system queries a driver for a destination. The navigation  
20 application may submit an interaction that asks for a destination state. If the state  
21 is submitted, the application may then submit an interaction that asks for a  
22 destination city. If the driver submits the city, the application may then submit an  
23 interaction that asks for the destination address.  
24  
25

1 It is easy to understand why a navigation application would not want these  
2 interactions broken up. If the interactions are separated, the driver or the speech  
3 system 124 may become confused as to where the other is in the dialogue.

4 Fig. 5 is a flow diagram depicting the methodology of interaction chaining.  
5 Similar to Fig. 4, a current state of the interaction list 168 is shown at each stage of  
6 the flow diagram. It is noted that, for this example, one of the applications 130 -  
7 134 submits a conversation to be processed. The conversation consists of  
8 interaction\_A 170 and interaction\_n 176.

9 At block 500, interaction\_A 170 is active while interaction\_B 172 and  
10 interaction\_C 174 are inactive and waiting in the interaction list 168 to be  
11 processed. After interaction\_A 170 concludes processing at block 502, the  
12 interaction manager 160 waits for the pre-determined grace period before moving  
13 on to processing interaction\_B 172 (block 504).

14 At block 606, the application that submitted interaction\_A 170 submits  
15 interaction\_n 176 to be processed to complete the conversation. The submission  
16 of interaction\_n 176 occurs before the grace period has expired. If interaction\_n  
17 176 is not submitted before the grace period expires, interaction\_B 172 will begin  
18 processing.

19 When interaction\_n 176 is submitted before the grace period expires ("Yes"  
20 branch, block 506), interaction\_n 176 is processed immediately at block 508.  
21 There are no additional interactions to be processed after interaction\_n 176 has  
22 completed processing ("No" branch, block 506), so interaction\_B 172 begins  
23 processing at block 510. The desired result is achieved, because the complete  
24 conversation (interaction\_A 170 and interaction\_n 176) was processed without  
25 separating the interactions.

1 Although it is not typically desired, chained interactions may be interrupted  
2 by another application. If an application submits an interaction that is flagged to  
3 be processed immediately, that interaction will be placed at the front of the  
4 interaction list 168, even if doing so will interrupt a conversation. This is one  
5 reason that use of the 'process immediately' option should be used sparingly by  
6 applications. An example of when the 'process immediately' option may be used  
7 is when an automobile engine is overheating. It is probably desirable to interrupt  
8 any interactions being processed to tell the driver of the situation since the  
9 situation requires immediate attention.

#### 10 Chained Interaction Interruption: Methodological Implementation

11 Fig. 6 is a flow diagram depicting the process of interrupting a chained  
12 interaction. Once again, a current state of the interaction list 168 is shown  
13 corresponding to each portion of the flow diagram. Also, it is assumed that an  
14 application wants to process a conversation consisting of interaction\_A 170 and  
15 interaction\_n 176.

16 At block 600, interaction\_A 170 is active while interaction\_B 172 and  
17 interaction\_C 174 are inactive and waiting in the interaction list 168 to be  
18 processed. When interaction\_A 170 concludes processing at block 702, a grace  
19 period is established at block 604.

20 If no interaction is submitted by the same application ("No" branch, block  
21 606), then interaction\_B 172 is processed at block 608. However, in this example,  
22 interaction\_n 176 is submitted before the grace period expires ("Yes" branch,  
23 block 606). Therefore, interaction\_n 176 begins processing at block 610. At  
24 block 612, interaction\_m 198 is submitted and is flagged to be processed  
25 immediately, so it begins processing at block 614. Interaction\_m 198 continues to

1 be processed until it is completed ("No" branch, block 616). When interaction\_m  
2 198 has concluded ("Yes" branch, block 616), the interaction manager 160  
3 determines if interaction\_n 176 (which was interrupted) is set to self-destruct in  
4 the event that it is interrupted. If interaction\_n 176 is to self-destruct ("Yes"  
5 branch, block 618), then interaction\_B 172 begins to be processed at block 608. If  
6 interaction\_n 176 does not self-destruct ("No" branch, block 618), then  
7 interaction\_n 176 finishes processing at block 620.

#### 8 Grace Period Interruption: Methodological Implementation

9 Interruptions may also occur during a grace period, because the grace  
10 period does not preclude any application from interrupting. Fig. 7 is a flow  
11 diagram that depicts the process that takes place when an application submits an  
12 interrupting interaction during a grace period. As before, a current state of the  
13 interaction list 168 is shown corresponding to the blocks of the flow diagram.

14 At block 700, interaction\_A 170 is active while interaction\_B 172 and  
15 interaction\_C 174 are inactive and waiting in the interaction list 168 to be  
16 processed. When interaction\_A 170 concludes processing at block 702, a grace  
17 period is established at block 704.

18 Before the grace period has timed out, interaction\_n 176 interrupts and is  
19 placed at the front of the interaction list 168 (block 708). It is noted that  
20 interaction\_n 176 is not a part of the conversation that began with interaction\_A  
21 170. Interaction\_n 176 is processed at block 708 for as long as the interaction  
22 needs to run ("No" branch, block 710). Only when interaction\_n 176 has  
23 concluded processing ("Yes" branch, block 710) will interaction\_B 172 - the  
24 second interaction of the conversation - be processed (block 712).

25 Do Not Add Interaction to Non-Empty List

1 An application may also indicate that an interaction is not to be added to the  
2 interaction list if the interaction list is not empty at the time the interaction is  
3 submitted. One scenario in which this might be desirable is in the event that an  
4 application included a verbal clock that announced a current time every minute.  
5 If, during the time where the minute would normally be announced, another  
6 application was speaking to the user, the announcement interaction would not be  
7 added to the interaction list, because the announcement might be out of date by the  
8 time it is processed.

9 Another scenario might be a navigation application that announces a  
10 current location, block by block, as one drives, e.g., "You are on 1<sup>st</sup> and Main" . . .  
11 "You are on 2<sup>nd</sup> and Main," etc. It would not be desirable to add such interactions  
12 to the interaction list if the driver were speaking to another application.

### 13 Exemplary Grammar(s) & Grammar Attributes

14 The interaction manager 160 must also use specific attributes of each  
15 grammar that it processes to process grammar interactions correctly. When the  
16 speech system 102 is initially booted, any applications that are present at startup  
17 are registered with the master grammar table 162 (whether running or not) so that  
18 the speech system 102 is aware of each grammar that may possibly be active.  
19 Additionally, if an application launches or is added while the speech system 102 is  
20 running, the application will register its grammar in the master grammar table 162.

21 Fig. 8a is an illustration of a master grammar table 800 similar to the master  
22 grammar table 162 shown in Fig. 1. The master grammar table 800 is a table of  
23 grammar tables, there being one grammar table for each grammar available to the  
24 system.  
25

1 As shown in Fig. 8a, a grammar table 802 for grammar\_1 136 is included in  
2 the master grammar table 800. Similarly, a grammar table 804 for grammar\_2 138  
3 , a grammar table 806 for grammar\_3 140 and a grammar table 808 for  
4 grammar\_4 152 are included in the master grammar table 800. It is noted that  
5 practically any number of grammar tables may be stored in the master grammar  
6 table 800 between grammar table 802 and grammar table 806.

7 Fig. 8b is a more detailed illustration of a grammar table 810 similar to the  
8 grammar tables 802 - 806 shown in Fig. 8a. Grammar table 810 includes several  
9 members: a grammar identifier 820; an executable command 822; a global flag  
10 826; a persistent flag 828; an active flag 830; and a static flag 832. Each of the  
11 members 820 - 832 included in the grammar table 810 specifies an attribute of a  
12 grammar associated with the grammar table 810.

13 The grammar identifier 820 is a value that is uniquely associated with a  
14 grammar that corresponds to the grammar table 810. The grammar identifier 820  
15 is used with interactions to identify a grammar that is associated with the grammar  
16 identifier. Including the grammar identifier 820 with an interaction solves a  
17 problem of latency that is inherent in the speech system 102. After an application  
18 submits an interaction that is placed in the interaction list 168 of the interaction  
19 manager 160, the application must wait until the interaction reaches the front of  
20 the interaction list 168 before it is processed. When the interaction finally reaches  
21 the front of the interaction list 168, the speech server 124 immediately knows  
22 which grammar from the master grammar table 162 is associated with and,  
23 therefore used with, the interaction. If the grammar identifier 820 were not  
24 included in the interaction, the speech server 124 would first have to notify the  
25 application that the interaction submitted by the application is about to be



1 processed. Then, the speech server 124 would have to wait for the application to  
2 tell it which grammar to utilize. Since the grammar identifier 820 is included with  
3 a submitted interaction, the speech server can begin processing the interaction  
4 immediately.

5 The executable command 822 is a command (including a path if necessary)  
6 that may be used to launch an application associated with the grammar table 820.  
7 This allows the speech server 124 to launch an application with the executable  
8 command 822 even though the corresponding application is not loaded into the  
9 system. If the speech server 124 receives an indication that a recognition occurs  
10 for a particular grammar, the speech server 124 passes the recognition to an  
11 application that has registered the grammar if such an application is running. If,  
12 however, no application using the identified grammar is running, the speech server  
13 124 launches the application and passes the recognition to the application. This  
14 solves the problem of having to first launch an application manually before it may  
15 receive a command.

16 The speech

17 For example, suppose an automobile driver is driving down the road when  
18 she decides she wants to play an MP3 file by, say, David Bowie, on the automobile  
19 radio. Assume for this example, that the executable command 822 is a typical  
20 path such as “\win\ . . \mp3.exe” and that the recognition term 824 is “play mp3.”

21 Instead of having to manually activate an MP3 player and then command it  
22 to “play David Bowie,” the driver simply commands the system to “play MP3  
23 David Bowie.” Even though the MP3 player may not be running, the speech  
24 server 124 will recognize the command “play MP3” and execute the executable  
25 command 822 to start the MP3 player. The grammar associated with the MP3

1 player (not shown) will recognize "David Bowie" and play the desired selection  
2 that is associated with that command.

3 The global flag 826 is a value that, when set, indicates that the grammar  
4 associated with the grammar table 810 is a global grammar that may not be  
5 interrupted by another application or the speech system 102 (but only the same  
6 application). If the global flag 826 is not set, then the grammar is a yielding  
7 grammar that can be interrupted by other applications or by the speech system  
8 102. As will be discussed in greater detail below, a global grammar is always  
9 active, although parts of it may be deactivated by the application to which it  
10 corresponds.

11 It is noted that the global flag 826 may be implemented as a yielding flag  
12 (not shown) which, when set, indicates that the grammar is not a global grammar.  
13 The logic described for utilizing the global flag 826 would, in that case, simply be  
14 reversed.

15 The persistent flag 828 is a value that, when set, indicates that the grammar  
16 associated with the grammar table 810 is persistent and not transient. A persistent  
17 grammar is a grammar that is loaded by default when the speech system 102 is  
18 running, irrespective of the run state of its corresponding application. If the  
19 persistent flag 828 is set, then the grammar associated with the grammar table  
20 should not be removed from the master grammar table 800.

21 The active flag 830 is a value that, when set, indicates that the grammar  
22 associated with the grammar table 810 is currently active. When a grammar is  
23 active, the speech recognitions system 102 actively listens for the commands  
24 included in the grammar. When an interaction is submitted to the interaction  
25 manager 160, the interaction manager 160 indicates to the speech server 124 that

1 other grammars should yield to a certain grammar if applicable. The speech server  
2 124 sets the active flag 830 to a value that indicates the grammar associated with  
3 the grammar table 810 is active. Simultaneously, the interaction manager 160 will  
4 clear the active flag 830 for each *yielding* grammar in the master grammar table  
5 162. As a result, the set of commands that the speech system 102 listens for is  
6 reduced.

7 When the yielding grammars are de-activated, i.e., the active flags are  
8 cleared, any grammar that is global (i.e., the global flag 826 is set) remains active.  
9 This is because a global grammar is always active. Therefore, at any given time  
10 that an application is executing, the speech system 102 is listening for all global  
11 grammars in the master grammar table 800 and one yielding grammar that is  
12 currently active (i.e., is associated with the application that is currently executing)  
13 in the master grammar table 800. If no application is currently executing, the  
14 speech system 102 listens for all grammars, whether global or yielding.

15 In one implementation, the speech server 124 does not de-activate all  
16 yielding grammars other than a grammar associated with a currently executing  
17 application unless an interaction in the interaction list 168 includes a method that  
18 informs the speech server 124 that all other yielding grammars should be de-  
19 activated. When the interaction manager 160 identifies such a method, the  
20 interaction manager 160 sends a message to the speech server 124 to de-activate  
21 all other yielding grammars in the master grammar table 162.

22 Finally, the static flag 832 is a value that, when set, indicates that the  
23 grammar associated with the grammar table 810 is a static grammar and, therefore,  
24 will not change after it is registered in the master grammar table 162.

## 25 Miscellaneous Functional Scenarios

1 The functional scenarios that follow are not discussed in detail with respect  
2 to the speech system 102, but may also be implemented with the features  
3 described above. The functional scenarios merely require that the interaction  
4 manager 160 be configured to handle the scenarios.

### 5 **Push-To-Talk**

6 Push-to-talk (PTT) is used to indicate that a command from the user is  
7 imminent, which allows a user to initiate a command. For example, a user may  
8 PTT and say "lock the doors" to actuate a vehicle's door locks. A push-to-talk  
9 (PTT) event instantiated by a user interrupts any current interaction.

10 PTT may also be used to provide a response to a system-initiated  
11 interaction. For example, if a navigation application asks "Fast or scenic route,"  
12 the user pushes push-to-talk and answers "fast" or "scenic."

### 13 **Barge-in**

14 The speech server 124 may also be configured to allow a user to "barge in"  
15 with a response. For example, if a navigation application asks "Fast or scenic  
16 route," the user may interrupt - without PTT - and answer "fast" or "scenic."

### 17 **Immediate Response to User Command**

18 The speech server 124 may be configured to provide an immediate response  
19 to a user command. For example, while an automobile system is announcing a  
20 driving instruction to a driver, the driver commands the system to "disconnect."  
21 The speech server 124 either disconnects immediately or confirms the disconnect  
22 command by stating "OK to disconnect", interrupting the original driving  
23 instruction.

### 24 **Application-aborted Interaction**

25

1 The applications 170 - 176 may also abort an interaction in certain  
2 circumstances. For example, a navigation application needs to tell a driver that a  
3 point of interest is drawing near, but other applications are currently talking to the  
4 driver. By the time the other applications have concluded, the point of interest is  
5 passed. The navigation application aborts the announcement interaction before it  
6 begins. If the point of interest has not been passed, the announcement is made,  
7 delaying only until the other applications have concluded.

### 8 **Interaction-specific Grammar**

9 The speech server 124 may also de-activate some grammars and leave  
10 active an interaction-specific grammar. For example, a navigation application asks  
11 a driver "fast or scenic route." Since the interaction is expecting a specific reply  
12 for a specific grammar, the specific grammar is activated (or remains active) to  
13 give the words "fast" and "scenic" priority over other grammars. This reduces the  
14 overhead required to process the driver's response, since the speech server 124  
15 does not have to listen for as many terms.

### 16 **Enhanced Prompt After Interruption**

17 The speech server 124 may also be configured to enhance a prompt during  
18 an interrupted conversation. If, for example, a navigation application asks for the  
19 driver's destination by stating first "please say the state." The driver responds  
20 with the destination state. The navigation application then asks "please say the  
21 city." However, during the announcement or before the driver answers with the  
22 destination city, the question is interrupted with an important announcement.  
23 After the announcement concludes, the original conversation resumes. To make  
24 up for the lost context, the speech server 124 is configured to revise the question to  
25 "for your destination, please say the city." By re-focusing the driver on the

1 navigation application conversation, the driver is less likely to be confused about  
2 what the system is saying.

### 3 Speech Server Interface

4 In order for the speech server 124 to support the grammars described above  
5 for the speech-enabled applications 130 - 134, the speech server 124 exposes  
6 several methods in the speech server interface 148 to the speech-enabled  
7 applications 130 - 134. Fig. 13 is a representation of the speech server interface  
8 148, 1300 and the methods 1302 - 1350 it exposes, which are described below.

9 In some cases, exemplary implementations are specific for use with the  
10 Speech Application Programming Interface (SAPI) provided by MICROSOFT  
11 CORP. It is not intended that these examples limit the speech server 124 or the  
12 speech server interface 148 to operation with the MICROSOFT CORP. SAPI.  
13 Those skilled in the art will recognize certain methods, functions, parameters, etc.,  
14 that are specific to the MICROSOFT CORP. SAPI, which may be altered to  
15 operate with other versions of the Speech API 126.

16 A **Create Grammar method** 1302 is used to load an existing grammar by  
17 name or create a new empty grammar. Flags may be used to determine different  
18 parameters about the grammar.

19 Exemplary implementation:

20 *CreateGrammar(WCHAR\* szFile, SPEECH\_LOAD\_OPTIONS*  
21 *LoadOptions SPEECH\_CONTEXT\_OPTIONS ContextOptions, DWORD*  
*\*pdwGrammarId)*

22 A **Get Grammar ID method** 1304 is used by an application with a  
23 persistent grammar to get a unique identifier assigned for that particular grammar.

24 Exemplary implementation:

25 *GetGrammarId(WCHAR\* szPersistenceId, DWORD\* pdwGrammarId)*

1       A **Remove Grammar method** 1306 removes a grammar from the speech  
2 server. The grammar is removed even if it is a persistent grammar.

3       Exemplary implementation:

4       *RemoveGrammar(DWORD dwGrammarId)*

5       A **Persist method** 1308 is called to persist the grammar. It is used when an  
6 application is creating a grammar in code and then wants it persisted.

7       Exemplary implementation:

8       *Persist(DWORD dwGrammarId, WCHAR\* szLaunchPath, WCHAR\**  
9 *szPersistenceId)*

10       An **Advise Speech Events method** 1310 lets the speech server know that  
11 an application is listening for speech recognition events. The application uses  
12 flags to specify events that it wants to listen to.

13       Exemplary implementation:

14       *AdviseSpeechEvents(DWORD dwSpeechFlags, DWORD dwReserved,*  
15 *OLE\_HANDLE hwndNotify)*

16       An **Unadvise Speech Events method** 1312 lets the speech server know  
17 that an application is no longer listening for the specified events.

18       Exemplary implementation:

19       *UnadviseSpeechEvents(DWORD dwSpeechFlags, DWORD dwReserved,*  
20 *OLE\_HANDLE hwndNotify)*

21       A **Yield To Grammar method** 1314 makes all yielding grammars yield,  
22 except for the grammar that is passed in. This method is used by the interaction  
23 manager 160 to restrict the grammars that are active during a conversation.

24       Exemplary implementation:

25       *YieldToGrammar(DWORD dwGrammarId)*

1       An **Unyield To Grammar method** 1316 unyields all yielding grammars,  
2 except for the grammar that is passed in. This method is used by the interaction  
3 manager 160 to restrict the grammars that are active during a conversation.

4       Exemplary implementation:

5       *UnyieldToGrammar(DWORD dwGrammarId)*

6       A **Commit method** 1318 is used to commit a grammar to the speech server  
7 and to the Speech API 126.

8       Exemplary implementation:

9       *Commit(DWORD dwGrammarId, DWORD dwReserved)*

10       A **Get Rule method** 1320 is used as a wrapper around the MICROSOFT  
11 CORP. SAPI ISpGrammarBuilder::GetRule method. The speech server uses this  
12 method to construct and control individual rules in a grammar.

13       Exemplary implementation:

14       *GetRule(DWORD dwGrammarId, WCHAR\* szRuleName, DWORD*  
15 *dwRuleId, DWORD dwAttributes, BOOL fCreateIfNotExist, l HANDLE\* phState)*

16       A **Create New State method** 1322 is used by the speech server to create  
17 new states in the SAPI grammar.

18       Exemplary implementation:

19       *CreateNewState(DWORD dwGrammarId, HANDLE hOriginalState,*  
20 *HANDLE\* phNewState)*

21       An **Add Word Transition method** 1324 adds a transition between two  
22 states on a word.

23       Exemplary implementation:

24       *AddWordTransition(DWORD dwGrammarID, HANDLE hFromState,*  
25 *HANDLE, hToState, WCHAR\* szPhrase, WCHAR\* szSeparators,*  
*SPEECH\_GRAMMAR\_WORD\_TYPE GrammarWordType, float fWeight,*  
*SPPROPERTYINFO \*pPropInfo)*



1 An **Add Rule Transition method** 1326 adds a transition between two rules  
2 on a word.

3 Exemplary implementation:

4 *AddRuleTransition(DWORD dwGrammarId, HANDLE hFrom, HANDLE*  
5 *hTo, HANDLE hRule, float fWeight, SPPROPERTYINFO \* pPropInfo)*

6 A **Set Rule State method** 1328 activates and de-activates rules.

7 Exemplary implementation:

8 *SetRuleState(DWORD dwGrammarId, WCHAR\* szRuleName, DWORD*  
9 *dwRuleId, SPRULESTATE NewState)*

10 A **Set Grammar State method** 1330 activates and de-activates grammars.

11 Exemplary implementation:

12 *SetGrammarState(DWORD dwGrammarId, SPGRAMMARSTATE*  
13 *GrammarState)*

14 A **Get Grammar State method** 1332 is used to get a grammar state.

15 Exemplary implementation:

16 *GetGrammarState(DWORD dwGrammarId, SPGRAMMARSTATE\**  
17 *pGrammarState)*

18 A **Get Récognition method** 1334 is called to get a recognition that has  
19 occurred.

20 Exemplary implementation:

21 *GetRecognition(DWORD dwGrammarId, DWORD dwRecold, BYTE*  
22 *\*\*ppSerializedPhrasel, DWORD \* pdwSize)*

23 A **Get Alternate method** 1336 is called to get alternates to a recognition  
24 that has occurred.

25 Exemplary implementation:

*GetAlternate(DWORD dwGrammarId, DWORD\* pdwAlternateCookie,*  
*DWORD dwRecold, BYTE \*\*ppSerializedPhrase, DWORD \* pdwSize)*

1       A **Turn Recognizer On method** 1338 is used to turn the speech recognizer  
2 122 on.

3       Exemplary implementation:

4       *TurnRecognizerOn(DWORD dwTimeout)*

5       A **Turn Recognizer Off method** 1340 is used to turn the speech recognizer  
6 122 off.

7       Exemplary implementation:

8       *TurnRecognizerOff(DWORD dwReserved)*

9       A **Get Recognizer State method** 1342 is used to get a speech recognizer  
10 122 state.

11       Exemplary implementation:

12       *GetRecognizerState(BOOL \*pval)*

13       An **Advise SAPI Event method** 1344 registers interest in SAPI 126 events.  
14 A sink that is passed in is called when an event that is advised for occurs.

15       Exemplary implementation:

16       *AdviseSAPIEvent(ISpeechEventSink\* pSink, ULONGLONG ullEvents)*

17       An **Unadvise SAPI Event method** 1346 lets the speech server 124 know  
18 that the sink is no longer interested in SAPI 126 events.

19       Exemplary implementation:

20       *UnadviseSAPIEvent(ISpeechEventSink\* pSink)*

21       A **Get Recognition Context method** 1348 gets a speech recognition  
22 context pointer from the speech engine 118.

23       Exemplary implementation:

24       *GetRecoContext(IUnknown\*\* ppunkRecoContext)*

25       A **Get Voice method** 1350 gets a voice pointer from the speech engine 118.

Exemplary implementation:

*GetVoice(IUnknown\*\* ppunk Voice)*

### Speech Controls

The speech controls 154 - 158 are provided for use in speech-related applications to provide timesaving tools to developers who create the applications to run with the speech server 124. The speech controls 154 - 158 are computer-executable code modules that provide standardized functions for developers to use for common interactions utilized in speech-enabled applications, thereby saving the developers the time and effort required to code the interaction for each use.

### **Question Control**

The question control 154 gives an application developer an easy way to create various modal, system-initiated interactions, or dialogues. Such interactions are used to obtain information from a user by asking the user a question. The following scenarios exemplify common uses of the question control to obtain desirable characteristics.

**User Interface Consistency:** A user tries an in-car computer system in his friend's car. He then goes out to shop for a new car. He notices that although other systems sound a little different, working with their speech user interface dialogues is just the same.

**Application Compatibility:** A user buys a full-featured navigation system software package for her car computer. She then buys a new car of a different make. She is still able to install her navigation software in her new car and it works the same as it did in her old car.

1 Hardware/Software Compatibility: A developer can design a unique  
2 speech hardware and/or software subsystem to work in conjunction with the  
3 question control without compromising application compatibility or user interface  
4 consistency.

5 The question control allows flexible programming so that a variety of  
6 question scenarios can be implemented. For example, the question control may be  
7 used to ask a driver a simple question that may be answered "yes" or "no", or a  
8 more complex question such as "fast or scenic route" and receive "fast" or  
9 "scenic" as appropriate answers.

10 The question control also allows greater flexibility by allowing the use of  
11 dynamic grammars. A question control has a grammar associated with it. In the  
12 above examples, the grammar may only consist of "yes" and "no" or "fast" or  
13 "scenic." The question control can be configured by a developer or OEM to  
14 standardize behavior of certain types of questions that can't be provided with a  
15 simple list. For example, a hierarchical grammar such as a time or date grammar  
16 may be associated with a question control. Such types of grammars involve too  
17 many list choices to practically list for a user.

18 The question control may also be used to provide an interrupting question.  
19 For example, while a system is reading a news story via TTS, a car application  
20 asks "<ding> - Your gas tank is close to empty; do you want instructions to the  
21 nearest gas station?" Similarly, a question programmed with the question control  
22 may be interrupted. For example, while an e-mail application is asking "You have  
23 mail; do you want to read it now?" a car application announces, "<ding> - Your  
24 engine is overheating."

25 Table 1 lists question control properties and types. Discussion follows.

PROPERTY	TYPE
Type	Enumeration
Interrupting	Boolean
Prompt	String
Prompt Verbose	String
Earcon Mode	Enumeration
App-Provided Grammar	Grammar
List Choices	Boolean
Selection Feedback	Enumeration

Table 1  
Question Control Properties

TYPE PROPERTY - The question control supports a Type property that can be used to determine the behavioral or content characteristics of the application using the question control. The Type property ultimately determines properties used in defining the application's behavior.

INTERRUPTING PROPERTY - The Interrupting property determines whether the application will interrupt other interactions in the interaction list 168 of the interaction manager 160. If the Interrupting property value is true, then the application (i.e., the question created with the question control) interrupts any other interaction in the interaction list 168. If the Interrupting property is false, then the application does not interrupt other interactions, but places its interactions at the end of the interaction list 168.

PROMPT PROPERTY - The question control is able to verbally prompt a user in order to solicit a response. The Prompt property contains what is announced when the application/question is started. The Prompt property value is interpreted according to the value of a PromptType property, which is text-to-speech or pre-recorded. If the prompt is TTS, then the prompt announces the TTS string. If the prompt is pre-recorded, then the prompt announces the contents of a file that contains the recording.

PROMPT VERBOSE PROPERTY - The Prompt Verbose property is a prompt that an interaction plays if the application/question is re-activated after it is interrupted. This property may be NULL and, if so, the interaction plays whatever is specified by the Prompt property (the prompt initially stated at the beginning of the interaction (i.e., application/question). Similar to the Prompt property, the Prompt Verbose property includes a PromptType that may be a TTS string or a string stored in a file.

EARCON MODE PROPERTY - The Earcon Mode property determines if the question control will play an audio file when the question control is activated or re-activated. The audio file played is determined by a currently selected Type property. The Type property may be "Always," "After Interruption" or "Never."

If the Type property is "Always," then the audio file always plays on activation or re-activation. For example, if the audio file is a "ding" then the "ding" will be played when the system initiates a sole interaction or a first interaction in a conversation.

If the Type property is "After Interruption," then the audio file is only played on re-activation. For example, if a car system asks a user "Fast or scenic

1 route" after first being interrupted by a global announcement, the audio file (i.e.,  
2 "ding") sounds before the question repeats after the interruption.

3 If the Type property is "Never," then the audio file is never played. The  
4 application may modify the Type property between "Always" and "Never." The  
5 "Never" Type property may be set by an application when the application has a  
6 special need not to play the audio file.

7 APPLICATION-PROVIDED GRAMMAR - An application can provide  
8 the question control with a list of options from which the user may choose. For  
9 each option offered, the application may provide one or more phrases whose  
10 recognition constitutes that choice. Any choices added are in addition to any  
11 grammars implemented in the question control. For example, a navigation  
12 application may provide a list having two options, "fast" and "scenic." If the  
13 words "fast" and "scenic" are not already included in an active grammar, then they  
14 are automatically added.

15 In one implementation, the question control provides a 'spoken choice'  
16 feature. The spoken choice feature may be used when a question is configured to  
17 have two or more possible answers for one answer choice. For example, a  
18 question may ask "What is the current season?" The answers may be "Spring,  
19 Summer, Autumn and Winter." In addition, the word "Fall" may be used instead  
20 of "Autumn." The question control may be configured to respond to a user  
21 inquiry as to possible answers as including either "Autumn" or "Fall." As a result,  
22 the list choices provided to a user would be "Spring, Summer, Autumn and  
23 Winter," or "Spring, Summer, Fall and Winter."

24 Another user for the spoken choice feature is for speech systems that may  
25 mispronounce one or more words. For example, many speech systems will

1 mispronounce Spokane, Washington as having a long “a” sound, since that is how  
2 phonetics rules dictate (instead of the correct short “a” sound). If a speech system  
3 is to announce the word “Spokane” to a user, the question control (or another  
4 control) can be programmed to play a designated audio file that correctly  
5 pronounces Spokane instead of using a standard TTS.

6 The application’s various grammars are activated in the system  
7 immediately upon starting the control. This provides for the user’s ability to barge  
8 in (using push-to-talk) and respond to the question control before it is finished.

9 LIST CHOICES PROPERTY - The List Choices property determines  
10 whether the question control will automatically TTS the list of valid choices to a  
11 user after playing the prompt. This option is particularly useful when the user is  
12 likely to be unaware of the valid responses. For example, a navigation application  
13 may ask a driver who has just entered a destination “Which route would you like  
14 to take, fast or scenic?”

15 SELECTION FEEDBACK PROPERTY - The Selection Feedback property  
16 determines if the question control will play feedback automatically when the user  
17 answers one of the application-provided or system-provided options that are  
18 enumerated by the List Choices property. If the Selection Feedback property has a  
19 value of “None,” no feedback is played when the user makes a choice. If the  
20 Selection Feedback property has a value of “Earcon,” then a designated  
21 satisfaction earcon is played when the user makes a choice. If the Selection  
22 Feedback property has a value of “Echo Choice” value, then a TTS of the user’s  
23 choice is played when the user makes a choice.



1 Fig. 10 is a flow diagram depicting a question control process. The  
2 question control process depicted in Fig. 10 is only one way in which the question  
3 control may be implemented.

4 At block 1000, the question control is launched. If there is an earcon to be  
5 played to indicate a question prompt is about to be asked ("Yes" branch, block  
6 1002), then the earcon is played at block 1004. Otherwise, no earcon is played  
7 ("No" branch, block 1002). The question prompt is then played at block 1008.

8 The choices with which the user may respond to the question prompt may  
9 be announced for the user at block 1010 ("Yes" branch, block 1008). But this may  
10 not be desirable and, therefore, the play list choices block may be skipped ("No"  
11 branch, block 1008).

12 Just as an earcon may be played to alert the user that a question prompt is  
13 forthcoming, an earcon may also be played after the question (block 1014) prompt  
14 to indicate to the user that the system is ready for the user's answer ("Yes" branch,  
15 block 1012). If this is not desirable, the application may be programmed so that  
16 no such earcon is played ("No" branch, block 1012).

17 Blocks 1016 - 1026 represent the possible user responses to the question  
18 prompt (block 1008). At block 1016, the user may answer "What can I say?"  
19 ("Yes" branch, block 1016) indicating that the user desires to hear the possible  
20 responses to the question prompt. Control of the process then returns to block  
21 1010, where the play list choice prompt is repeated to the user.

22 If the user's response is to repeat the question prompt ("Yes" branch, block  
23 1018), then control of the process returns to block 1006, where the question  
24 prompt is repeated to the user. If the user's response is ambiguous, i.e., it is a  
25 response that the system does not understand ("Yes" branch, block 1020), then the

1 system TTS's "Answer is ambiguous" at block 1021. Control of the process  
2 returns to block 1012 to receive a new answer from the user.

3 If the question control receives a valid response from the user ("Yes"  
4 branch, block 1022), then feedback may be returned to the user to verify that the  
5 user has returned a valid response. If there is no feedback ("None" branch, block  
6 1034), then the result, i.e., the user's choice, is returned by the question control at  
7 block 1038. If the feedback is an earcon to indicate a valid response ("EC"  
8 branch, block 1034), then the earcon is played at block 1036 and the result is  
9 returned to the application at block 1038. If the feedback is to play TTS of the  
10 user's choice ("Echo" branch, block 1034), then the user's response is TTS'd to  
11 the user at block 1040 and the response is returned by the question control to the  
12 application at block 1038.

13 In one implementation of the question control described herein, a user may  
14 have an option to cancel a question process. If the user's response to the question  
15 prompt is to cancel ("Yes" branch, block 1024), and if canceled is enabled ("Yes"  
16 branch, block 1044), then the question is canceled. If an earcon is to be played to  
17 verify the cancellation ("Yes" branch, block 1046) then the appropriate earcon is  
18 played at block 1048 and a 'cancel' value is returned to the application to indicate  
19 the cancellation. If an earcon is not to be played upon cancellation ("No" branch,  
20 block 1046, then 'cancel' is returned at block 1050 without playing an earcon.

21 If the cancel option is not enabled ("No" branch, block 1044), then the  
22 system does not respond to the "cancel" command. If after a pre-determined  
23 timeout period elapses without receiving a response from the user ("Yes" branch,  
24 block 1026), the 'cancel' is returned to the application at block 1050. 'Cancel' is  
25 returned after an earcon is played (block 1048) if a cancel earcon is enabled ("Yes"

1 branch, block 1044). Otherwise ("No" branch, block 1048), 'cancel' is returned  
2 without first playing a cancel earcon. (Note that there is not a "No" branch to  
3 block 1026; this is due to the fact that if a response is returned, the response will  
4 have been handled before a determination is made as to whether a response was  
5 received during the timeout period.) Other implementations may handle the  
6 process of the control differently.

### 7 **Announcer Control**

8 The announcer control 155 provides a developer an easy way to deliver  
9 verbal feedback to users, including short notices and long passages of text-to-  
10 speech. The announcer control 155 implements a simple mechanism for playing  
11 pre-recorded speech or TTS text, and for giving a user standardized control of  
12 such playback. Use of the announcer control 155 significantly decreases the effort  
13 required by application developers to build a rich application user interface.

14 The following scenarios exemplify common applications of the announcer  
15 control 155.

16 READ E-MAIL: A user request that an electronic mail message be read.  
17 The system begins TTS'ing the message. The user is able to pause, fast forward,  
18 rewind, etc.

19 INTERRUPTING ANNOUNCER: While a navigation application is  
20 asking "Fast or scenic route?" the user commands "Read e-mail." The system  
21 begins to read the e-mail immediately.

22 INTERRUPTED ANNOUNCER: While the system is reading a news  
23 story via TTS, an automobile application asks "<ding> Your gas tank is close to  
24 empty. Do you want instructions to the nearest gas station?"  
25

1 NOTIFICATION: E-mail arrives while a user is driving and the system  
2 announces, "<ding> E-mail has arrived."

3 CONVERSATION STATEMENT: A user answers the last question to  
4 specify a navigation destination and the system announces, "Turn right at the next  
5 intersection."

6 REPEATED ANNOUNCEMENT: A navigation application announces,  
7 "<ding> Turn right at the next intersection." But the user did not hear it. The user  
8 says, "Repeat" and the system repeats the announcement.

9 The following features, or properties, may be available on the announcer  
10 control 155. Table 2 lists announcer control properties and types. Discussion  
11 follows.

PROPERTY	TYPE
Type	Enumeration
Interrupting	Boolean
ConversationID	String
Abort When Interrupted	Boolean
Earcon Mode	Enumeration
Announcement	String
Cancel Feedback	Boolean
Post Delay	Integer

Table 2  
Announcer Control Properties

TYPE PROPERTY: The announcer control 155 supports the Type property that can be used to determine the behavioral or content characteristics of the application/announcement. The Type property ultimately determines the properties used in defining the application's/announcement's behavior. The speech server 124 defines the Type property's valid values.

INTERRUPTING PROPERTY: The Interrupting property determines whether the application/announcement will interrupt other interactions present in the interaction list 168 of the interaction manager 160. If the Interrupting property value is True, an announcement interaction will immediately interrupt any other interactions in the interaction list 168. If the value is False, an announcement interaction will be placed at the end of the interaction list 168.

1 CONVERSATION ID PROPERTY: The Conversation ID property  
2 determines whether the application/announcement will operate in the context of  
3 the named conversation. The Conversation ID property is a string associated with  
4 a control instance. The interaction queue uses the Conversation ID property o  
5 identify which interaction belongs with which conversation.

6 ABORT WHEN INTERRUPTED PROPERTY: The Abort When  
7 Interrupted property determines whether the announcement will automatically  
8 self-destruct if it is interrupted by another interaction. If the property value is  
9 True, then the announcement aborts when interrupted; if the value if False, the  
10 announcement does not abort.

11 EARCON MODE Property: The Earcon Mode property determines if the  
12 application will play an audio file when it is activated or re-activated. If the  
13 Earcon Mode property has a value of "Always" the designated audio file is always  
14 played upon activation or re-activation. If the value is "After Interruption" the  
15 audio file is only played on re-activation; not on activation. If the value is  
16 "Never" an audio file is not played on activation or re-activation.

17 ANNOUNCEMENT PROPERTY: The Announcement property contains  
18 what is announced when the control is started. If an Announcement Type  
19 associated with the Announcement property is "TTS," then the Announcement  
20 property contains a string that is to be TTS'ed. If the Announcement Type is "Pre-  
21 recorded," then the Announcement property contains a string designating a file to  
22 be announced, i.e., a file name. If the Announcement Type is "Combination," then  
23 the Announcement property contains a TTS string and an audio file name.

24 CANCEL EARCON PROPERTY: The Cancel Earcon property determines  
25 if the announcer control will play an audio file automatically when the user

1 answers "cancel" (or its equivalent). If the Cancel Earcon property is True, then  
2 an earcon is played upon canceling; otherwise, an earcon is not played.

3 POST DELAY PROPERTY: The Post Delay property determines if the  
4 application will pause for a definable period of time after the announcement has  
5 been completely delivered. This features gives a user some time to issue a  
6 "repeat" or "rewind" command. It also provides for a natural pause between  
7 interactions. If the Post Delay property value is True, then a post delay is provided  
8 when not in the context of a conversation. If the value is False, then a post delay  
9 is not provided.

10 Fig. 11 is a flow diagram depicting an announcer control process. At block  
11 1100, the announcer control is activated at some time other than after an  
12 interruption. If an earcon mode associated with the announcer control that may be  
13 set to "Always," "Never," or "After Interruption." If the earcon mode is set to  
14 "Always" ("Always" branch, block 1102), then an earcon is played at block 1108,  
15 prior to an earcon being played at block 1108. If the earcon mode is set to  
16 "Never" or "After Interruption" mode ("Never or After Interruption" branch,  
17 block 1102), then an earcon is not played before an announcement is played at  
18 block 1108.

19 There may be a post delay after the announcement has completed ("Yes"  
20 branch, block 1112. If the user asks the system to repeat the announcement during  
21 a post delay period ("Yes" branch, block 1114), then the announcement is replayed  
22 at block 1110. If the user does not ask the system to repeat the announcement  
23 during the post delay period ("No" branch, block 1114), then the process  
24 completes at block 1116.  
25

1 A post delay may not be activated for the announcement control. If not  
2 (“No” branch, block 1112), then the process completes at block 1116 immediately  
3 after the announcement is played at block 1110.

4 Activation of the announcement control may occur after an interruption at  
5 block 1104. If an interruption occurs before the announcement control is activated  
6 and the announcement control earcon mode is set to play an earcon “Always” or  
7 “After Interruption” (“Always or After Interruption” branch, block 1106), then an  
8 earcon is played at block 1108 to alert the user that an announcement is  
9 forthcoming. The announcement is then played at block 1110. If the earcon mode  
10 is set to “Never” (“Never” branch, block 1106), then the announcement is played  
11 at block 1110 without playing an earcon at block 1108.

12 Thereafter, a post delay may be implemented (“Yes” branch, block 1112)  
13 wherein the user may ask the system to repeat the announcement (“Yes” branch,  
14 block 1114), in which case the announcement is repeated at block 1110. If a post  
15 delay is not implemented (“No” branch, block 1112), or if no response is received  
16 during a post delay period (“No” branch, block 1114), then the process concludes  
17 at block 1106.

## 18 Command Control

19 The command control 156 is designed to easily attach command-and-  
20 control grammar to an application. The command control 156 is used for user-  
21 initiated speech. At a minimum, the command control 156 must perform two  
22 functions. First, the command control 156 must provide a way for an application  
23 to specify what grammar(s) the application is interested in listening to. Second,  
24 the command control 156 must communicate back to the application that a  
25



1 recognition has occurred. To accomplish these tasks, the command control 156 is  
2 made up of four objects.

3 Fig. 12 is a block diagram of a command control 1200 similar to the  
4 command control 156 shown in Fig. 1. The command control 1200 includes a  
5 command manager object 1202, a grammar object 1204, a rule object 1206 and a  
6 recognition object 1208. For purposes of further discussion, the command control  
7 1200 is assumed to be an ActiveX control that conforms to ActiveX standards  
8 promulgated by Microsoft Corporation.

9 Each of the four objects 1202 - 1208 includes an interface: the command  
10 manager object interface 1210, the grammar object interface 1212, the rule object  
11 interface 1214 and the recognition object interface 1216. The interfaces 1210 -  
12 1216 of each object 1202 -1208 will be discussed separately in greater detail.

13 The command manager object interface 1210 has three properties:  
14 Persistence ID 1220; Grammar ID 1222; and Grammar 1224. The Persistence ID  
15 1220 is used to identify the application for persistence purposes. The Persistence  
16 ID 1220 must be unique in the system. The Persistence ID 1220 may be blank if  
17 the associated grammar is no persistent. In one implementation, the Persistence  
18 ID 1220 is a ProgID (Microsoft WINDOWS implementation).

19 The Grammar ID 1222 is an identifier that is used by with interactions 170  
20 -176 submitted to the interaction manager 160. As previously explained, the  
21 Grammar ID 1222 is utilized to avoid latency problems inherent in the speech  
22 system 102. The Grammar 1224 property is a pointer to the Grammar Object  
23 1204 that is associated with the Command Control 1200.

24 The command manager object interface also includes several methods:  
25 Create Grammar 1226, Persist 1228, Remove Grammar 1230, Start 1232 and

1 Event: Recognition 1234. Create Grammar 1226 is a function that is used to  
2 create a new grammar object from a grammar file. A grammar file may be an  
3 XML (extended markup language) file or a compiled grammar file (.cfg) or  
4 NULL, indicating that a new grammar is to be built. Parameters for Create  
5 Grammar 1226 include a path of a file to be opened or NULL for a new grammar  
6 (file), a value that indicates whether a grammar is static or dynamic (Load  
7 Options), a value that indicates whether a grammar is yielding or global (Context  
8 Options), and a pointer that receives the grammar object (ppGrammar).

9 Persist 1228 is a method that indicates that a grammar is to be persisted.  
10 Persisted grammars recognize even if the application with which they are  
11 associated are not running. If a recognition occurs, the application is launched.  
12 Persist 1228 includes two parameters: the grammar under which the ID should be  
13 persisted (Persistence ID); and a complete path for an executable that will handle  
14 grammar recognitions (Application Path).

15 Remove Grammar 1230 is a method that removes a grammar from the  
16 speech server 124. If the grammar is persistent, Remove Grammar 1230 un-  
17 persists the grammar. Start 1232 is a method that is called to let the speech server  
18 124 know that an application is ready to start handling events. Event: Recognition  
19 is a method that is called by the speech server 124 when a speech recognition  
20 occurs so that an appropriate application may be so notified.

21 A specific implementation of the command manager object interface 1210  
22 is shown below. The implementation is specific to the WINDOWS family of  
23 operating systems by Microsoft Corp. Other interfaces may be added to make the  
24 command control and ActiveX control (provided by the ATL wizard) so that a  
25 developer can simply drop the control on a form and proceed.

```

1 interface ICommandManager : IUnknown, IDispatch
2 {
3     Properties :
4         BSTR PersistenceID; (get/put)
5         DWORD GrammarID; (get only)
6         IDispatch* Grammar; (get only)
7
8     Methods :
9     CreateGrammar (BSTR File, SPEECH_LOAD_OPTIONS
10        LoadOptions, SPEECH_CONTEXT_OPTIONS
11        ContextOptions, IDispatch** ppGrammar)
12
13        HRESULT Persist (BSTR PersistenceID, BSTR
14        ApplicationPath)
15
16        HRESULT RemoveGrammar ()
17        HRESULT Start();
18 };
19
20 interface _ICommandManagerEvents: IDispatch // this
21     interface is the event that is sent back on recognition//
22
23 {
24     HRESULT Recognition(IDispatch * Recognition,
25         DWORD CountAlternates);
26 }

```

1 The Grammar Object Interface 1212 has an Enabled property 1236, a Rule  
2 method 1238, a Create Rule method 1240, and a Commit method 1241. The  
3 Enabled property 1242 is used to turn the entire grammar on or off. The Rule  
4 method 1248 selects a rule (by ID or name) and returns it to the caller. The Rule  
5 method 1248 includes a *RuleID* parameter that is either a numeric ID for the rule  
6 or a string for the rule name.

7 The Create Rule method 1240 creates a new rule in the grammar. The  
8 Create Rule method 1240 also utilizes the *RuleID* parameter, which is a name or  
9 numeric identifier of the rule to be created. Other parameters used in the Create  
10 Rule method 1240 include *Rule Level*, *Rule State*, *ppRule* and *Prop*. *Rule Level* is  
11 an enumeration determines whether the rule is created as a top level rule or not.  
12 *Rule State* specifies whether the rule is to be created as dynamic. Dynamic rules  
13 can be modified after they are committed. *ppRule* is the rule object that is created.  
14 *Prop* is an optional PropID or PropName that a developer wants to associate with  
15 the rule.

16 The Commit method 1241 method commits all changes made in the  
17 grammar and all of the rules.

18 A specific implementation of the grammar object interface 1212 is shown  
19 below. As with the command manager object interface shown above, the  
20 implementation is specific to the WINDOWS family of operating systems by  
21 Microsoft Corp.

```

1 interface IGrammar : IUnknown, IDispatch
2 {
3     Properties :
4         VARIANT_BOOL Enabled (get/put)
5
6     Methods :
7         IDispatch * Rule(VARIANT RuleID) (get only)
8         HRESULT CreateRule ([in] VARIANT RuleID,
9             SPEECH_RULE_LEVEL RuleLevel, SPEECH_RULE_STATE
10            RuleState, [out, retval] IDispatch **ppRule, [in, optional]
11            VARIANT Prop)
12
13         HRESULT Commit();
14 };

```

### Rule Class Interface

The Rule Class interface 1214 includes an enabled 1242 property and several methods: Add Rule 1244, Add Phrase 1246, Add Alternate Rule 1248 and Add Alternate Phrase 1250. Enabled 1242, when set, indicates whether a rule is active or inactive. Add Rule 1244 appends a rule to an existing rule structure. For example, if the rule looks like “Rule → Phrase Rule1” and Rule2 is added, then a new structure results, “Rule → Phrase Rule1 Rule2”.

In the WINDOWS specific implementation shown below, Add Rule 1244 includes two parameters, *plrule*, which is a pointer to the rule object that will be added to the rule. *Prop* is an optional PROPID or PROPNAME that can be associated with the rule.

Add Phrase 1246 appends a phrase to an existing rule structure. In the implementation shown below, the Add Phrase 124 method includes parameters *text* and *val*. *Text* is the text that is to be added. *Val* is an optional *val* or *valstr* that may be associated with the phrase. For this to be set, the rule must have been created with a property.

1       Add Alternate Rule 1248 places a new rule as an optional path for the  
2 previous rule structure. For example, if the structure is "Rule → Phrase Rule1"  
3 and then add alternative rule2 results in the new structure "Rule → (Phrase Rule1  
4 | Rule2. Concatenation takes precedence over the 'or' operator. Add Alternate  
5 Rule 1248 includes two parameters in the WINDOWS implementation shown  
6 below. *plrule* is a pointer to the rule object that will be added to the rule. *prop* is  
7 an optional PROPID or PROPNAME that may be associated with the rule.

8       Add Alternate Phrase 1250 places a new string as an optional path for the  
9 previous rule structure. If the structure is "Rule → (Phrase Rule1)" and alternative  
10 phrase Phrase2 is added, the new structure is "Rule → (Phrase Rule1) | Rule2.  
11 Concatenation takes precedence over the 'or' operator. In the WINDOWS  
12 implementation shown below, Add Alternate Phrase 1250 includes two  
13 parameters. *Text* is the text to be added. *Val* is an optional VAL or VALSTR that  
14 is to be associated with the phrase. The rule must have been created with a  
15 property for this to be set.

16       A specific implementation of the Rule Object interface 1214 is shown  
17 below. As with the other interfaces shown above, the implementation is specific  
18 to the WINDOWS family of operating systems by Microsoft Corp.

```
19       interface IRule : IUnknown, IDispatch
20       {
21       Properties :
22            VARIANT_BOOL Enabled (put only)
23       Methods :
24            HRESULT AddRule ([in] IDispatch *piRule, [optional, in]
25                VARIANT Prop)
26            HRESULT AddPhrase ([in] BSTR Text, [optional, in] VARIANT
27                Val)
28            HRESULT AddAlternative Rule ([in] IDispatch *piRule, [optional, in]
29                VARIANT Prop)
```

in] VARIANT Prop)

HRESULT AddAlternativePhrase ([in] BSTR Text, [optional, in]  
VARIANT Val)

};

EXAMPLE:

The Rule Object interface 1214 is designed for building grammars in a BNF (Backus-Naur Format) format. The rule is composed of a Start component that is constructed of either rules or phrases. The Start component corresponds to a top-level rule. For example:

$S \rightarrow A B \mid C$

$A \rightarrow \text{"I like"}$

$B \rightarrow \text{"Candy"} \mid \text{"Food"}$

$C \rightarrow \text{"Orange is a great color"}$

There are four rules here (S, A, B, C). There are four phrases: "I like"; "Candy"; "Food"; and "Orange is a great color." This grammar allows three phrases to be said by the user "I like candy," "I like food," or "Orange is a great color." To construct this, assume four rules have been created by a grammar object and then build the rules.

S.AddRule(A)

S.AddRule(B)

S.AddAlternativeRule(C)

A.AddPhrase("I like")

B.AddPhrase("Candy")

B.AddAlternativePhrase("Food")

C.AddPhrase("Orange is a great color.")

## Word Trainer Control

The word trainer control 158 provides an easy way to implement a speech-oriented work-training interaction with a user, in support of tasks that involve voice tags, such as speed-dial entries or radio station names. The entire word training process is implemented with a combination of the word trainer control and other GUI (graphical user interface) or SUI (speech user interface) controls. The word trainer primarily focuses on the process of adding the user's way of saying a phrase or verbally referencing an object in the recognizer's lexicon.

It is noted that the Word Trainer control 158 wraps the word trainer API (application programming interface) provided by MICROSOFT CORP. The features discussed below are available on the word trainer control 158.

An example of a functional scenario for the word trainer control is a user initiating voice tag training to complete creating a speed-dial entry for "Mom." The system prompts the user to say the name of the called party. The user responds, "Mom." Training is then complete.

Another example of a functional scenario for the word trainer control is a user who wants to place a call via voice command, but cannot remember the voice tag that was previously trained. The system helps the user using a question control: "Choose who you'd like to call by repeating the name. <Mom.wav>, <Dad.wav> or <work.wav>."

The following Tables (Tables 3-5) illustrate possible word training sessions that are supported by the word training control 158.



WHO	WHAT	DETAIL
System	Prompt	"Say name twice; Please say name"
System	Earcon	Signals user to start utterance
System	AutoPTT	Lets user talk w/o manual PTT
User	Utterance	Says "Mom"
System	Feedback	Plays <Mom.wav>
System	Prompt	"Please say the name again"
System	Earcon	Signals user to start utterance
System	AutoPTT	Lets user talk w/o manual PTT
User	Utterance	Says "Mom"
System	Feedback	Plays <Mom.wav>
System	Question	"OK to continue?"
System	Announcement	"You can now dial by saying <Mom.wav>

Table 3  
Scenario "A"

WHO	WHAT	DETAIL
System	Prompt	"Please say name"
User	PTT	User pushes PTT
System	Earcon	Signals PTT pushed, ready to record
User	Utterance	Says "Mom"
System	Earcon	Signals recording successful

Table 4  
Scenario "B"

WHO	WHAT	DETAIL
System	GUI Dialogue	Includes buttons for two training passes
User	Pushes #1	Starts training pass #1
System	Earcon	Signals PTT; Ready to record
System	AutoPTT	Lets user talk w/o manual PTT
User	Utterance	Says "Mom"
System	Feedback	Plays .wav of "Mom"
System	Disables #1	Shows that pass #1 remains
User	Pushes #1	Starts training pass #2
System	Earcon	Signals PTT; Ready to record
System	AutoPTT	Lets user talk w/o manual PTT
User	Utterance	Says "Mom"
System	Feedback	Plays .wav of "Mom"
System	Disables #2	Shows that pass #2 remains
System	GUI Dialogue	"Voice tag created"

Table 5  
Scenario "C"

Word Trainer is a control, such as an ActiveX control, that a developer can include in an application for the purpose of initiating and managing a training user interface process. All of the interfaces exposed by the Word Trainer API (MICROSOFT CORP.)

Table 6 identifies word trainer control 158 properties. It is noted that these properties are in addition to Word Trainer API (MICROSOFT CORP.) properties and methods wrapped by the word trainer control 158.

PROPERTY	TYPE
Type	Enumeration
Interrupting	Boolean
Feedback	Enumeration
PassesRemaining	Integer

Table 6

### Word Trainer Control Properties

The word trainer control 158 supports the Type property that can be used to determine the behavioral or content characteristics of the control. It is noted that it is the Type property that ultimately determines the style class and properties used in defining the control's behavior. The Type property's valid values are defined in the system's current speech theme.

The Interrupting property determines whether the control will interrupt other interactions in the interaction list 168 of the interaction manager 160. If the Interrupting property has a value of "True," then the control immediately interrupts any other interaction in the interaction list 168. If the value is "False," then the control does not interrupt, but places interactions at the end of the interaction list 168.

The Feedback property determines if the word trainer control 158 will play feedback automatically after the system successfully records the user. If the Feedback property has no value (or a value of 'none'), then the word trainer

1 control 158 doesn't play feedback when the user makes a choice. If the Feedback  
2 property has a value of "Earcon," then the word trainer control 158 plays a  
3 completion earcon resource after a successful recording. If the value is "Echo  
4 recording," then the word trainer control 158 plays a sound file of the user's  
5 recording.

6 The PassesRemaining property is a read-only property that tells an  
7 application how many recording passes the engine requires before a usable voice  
8 tag exists. It is intended that, as this number decrements, the application user  
9 interface reflects course progress through the training process.

10 In addition to the foregoing, the word trainer control 158 includes a  
11 StartRecording method. The StartRecording method initiates the recording  
12 process for one pass. When recording completes successfully, the  
13 PassesRemaining property decrements. It is noted that, in the cases where the  
14 speech engine can accept additional recordings, an application may call  
15 StartRecording even though PassesRemaining equals zero.

16 It is noted that other speech recognition grammars must be temporarily  
17 disabled when the speech engine is in a recording mode.

## 18 19 EXEMPLARY COMPUTER ENVIRONMENT

20 The various components and functionality described herein are  
21 implemented with a number of individual computers. Fig. 9 shows components of  
22 typical example of such a computer, referred by to reference numeral 900. The  
23 components shown in Fig. 9 are only examples, and are not intended to suggest  
24 any limitation as to the scope of the functionality of the invention; the invention is  
25 not necessarily dependent on the features shown in Fig. 9.

1 Generally, various different general purpose or special purpose computing  
2 system configurations can be used. Examples of well known computing systems,  
3 environments, and/or configurations that may be suitable for use with the  
4 invention include, but are not limited to, personal computers, server computers,  
5 hand-held or laptop devices, multiprocessor systems, microprocessor-based  
6 systems, set top boxes, programmable consumer electronics, network PCs,  
7 minicomputers, mainframe computers, distributed computing environments that  
8 include any of the above systems or devices, and the like.

9 The functionality of the computers is embodied in many cases by  
10 computer-executable instructions, such as program modules, that are executed by  
11 the computers. Generally, program modules include routines, programs, objects,  
12 components, data structures, etc. that perform particular tasks or implement  
13 particular abstract data types. Tasks might also be performed by remote  
14 processing devices that are linked through a communications network. In a  
15 distributed computing environment, program modules may be located in both local  
16 and remote computer storage media.

17 The instructions and/or program modules are stored at different times in the  
18 various computer-readable media that are either part of the computer or that can be  
19 read by the computer. Programs are typically distributed, for example, on floppy  
20 disks, CD-ROMs, DVD, or some form of communication media such as a  
21 modulated signal. From there, they are installed or loaded into the secondary  
22 memory of a computer. At execution, they are loaded at least partially into the  
23 computer's primary electronic memory. The invention described herein includes  
24 these and other various types of computer-readable media when such media  
25 contain instructions programs, and/or modules for implementing the steps

1 described below in conjunction with a microprocessor or other data processors.  
2 The invention also includes the computer itself when programmed according to  
3 the methods and techniques described below.

4 For purposes of illustration, programs and other executable program  
5 components such as the operating system are illustrated herein as discrete blocks,  
6 although it is recognized that such programs and components reside at various  
7 times in different storage components of the computer, and are executed by the  
8 data processor(s) of the computer.

9 With reference to Fig. 9, the components of computer 900 may include, but  
10 are not limited to, a processing unit 920, a system memory 930, and a system bus  
11 921 that couples various system components including the system memory to the  
12 processing unit 920. The system bus 921 may be any of several types of bus  
13 structures including a memory bus or memory controller, a peripheral bus, and a  
14 local bus using any of a variety of bus architectures. By way of example, and not  
15 limitation, such architectures include Industry Standard Architecture (ISA) bus,  
16 Micro Channel Architecture (MCA) bus, Enhanced ISA (EISA) bus, Video  
17 Electronics Standards Association (VESA) local bus, and Peripheral Component  
18 Interconnect (PCI) bus also known as the Mezzanine bus.

19 Computer 900 typically includes a variety of computer-readable media.  
20 Computer-readable media can be any available media that can be accessed by  
21 computer 900 and includes both volatile and nonvolatile media, removable and  
22 non-removable media. By way of example, and not limitation, computer-readable  
23 media may comprise computer storage media and communication media.  
24 "Computer storage media" includes both volatile and nonvolatile, removable and  
25 non-removable media implemented in any method or technology for storage of

1 information such as computer-readable instructions, data structures, program  
2 modules, or other data. Computer storage media includes, but is not limited to,  
3 RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM,  
4 digital versatile disks (DVD) or other optical disk storage, magnetic cassettes,  
5 magnetic tape, magnetic disk storage or other magnetic storage devices, or any  
6 other medium which can be used to store the desired information and which can be  
7 accessed by computer 910. Communication media typically embodies computer-  
8 readable instructions, data structures, program modules or other data in a  
9 modulated data signal such as a carrier wave or other transport mechanism and  
10 includes any information delivery media. The term "modulated data signal"  
11 means a signal that has one or more of its characteristics set or changed in such a  
12 manner as to encode information in the signal. By way of example, and not  
13 limitation, communication media includes wired media such as a wired network or  
14 direct-wired connection and wireless media such as acoustic, RF, infrared and  
15 other wireless media. Combinations of any of the above should also be included  
16 within the scope of computer readable media.

17 The system memory 930 includes computer storage media in the form of  
18 volatile and/or nonvolatile memory such as read only memory (ROM) 931 and  
19 random access memory (RAM) 932. A basic input/output system 933 (BIOS),  
20 containing the basic routines that help to transfer information between elements  
21 within computer 900, such as during start-up, is typically stored in ROM 931.  
22 RAM 932 typically contains data and/or program modules that are immediately  
23 accessible to and/or presently being operated on by processing unit 920. By way  
24 of example, and not limitation, Fig. 9 illustrates operating system 934, application  
25 programs 935, other program modules 936, and program data 937.



1 The computer 900 may also include other removable/non-removable,  
2 volatile/nonvolatile computer storage media. By way of example only, Fig. 9  
3 illustrates a hard disk drive 941 that reads from or writes to non-removable,  
4 nonvolatile magnetic media, a magnetic disk drive 951 that reads from or writes to  
5 a removable, nonvolatile magnetic disk 952, and an optical disk drive 955 that  
6 reads from or writes to a removable, nonvolatile optical disk 956 such as a CD  
7 ROM or other optical media. Other removable/non-removable,  
8 volatile/nonvolatile computer storage media that can be used in the exemplary  
9 operating environment include, but are not limited to, magnetic tape cassettes,  
10 flash memory cards, digital versatile disks, digital video tape, solid state RAM,  
11 solid state ROM, and the like. The hard disk drive 941 is typically connected to  
12 the system bus 921 through a non-removable memory interface such as interface  
13 940, and magnetic disk drive 951 and optical disk drive 955 are typically  
14 connected to the system bus 921 by a removable memory interface such as  
15 interface 950.

16 The drives and their associated computer storage media discussed above  
17 and illustrated in Fig. 9 provide storage of computer-readable instructions, data  
18 structures, program modules, and other data for computer 900. In Fig. 9, for  
19 example, hard disk drive 941 is illustrated as storing operating system 944,  
20 application programs 945, other program modules 946, and program data 947.  
21 Note that these components can either be the same as or different from operating  
22 system 934, application programs 935, other program modules 936, and program  
23 data 937. Operating system 944, application programs 945, other program  
24 modules 946, and program data 947 are given different numbers here to illustrate  
25 that, at a minimum, they are different copies. A user may enter commands and

1 information into the computer 900 through input devices such as a keyboard 962  
2 and pointing device 961, commonly referred to as a mouse, trackball, or touch  
3 pad. Other input devices (not shown) may include a microphone, joystick, game  
4 pad, satellite dish, scanner, or the like. These and other input devices are often  
5 connected to the processing unit 920 through a user input interface 960 that is  
6 coupled to the system bus, but may be connected by other interface and bus  
7 structures, such as a parallel port, game port, or a universal serial bus (USB). A  
8 monitor 991 or other type of display device is also connected to the system bus  
9 921 via an interface, such as a video interface 990. In addition to the monitor,  
10 computers may also include other peripheral output devices such as speakers 997  
11 and printer 996, which may be connected through an output peripheral interface  
12 995.

13 The computer may operate in a networked environment using logical  
14 connections to one or more remote computers, such as a remote computer 980.  
15 The remote computer 980 may be a personal computer, a server, a router, a  
16 network PC, a peer device or other common network node, and typically includes  
17 many or all of the elements described above relative to computer 900, although  
18 only a memory storage device 981 has been illustrated in Fig. 9. The logical  
19 connections depicted in Fig. 9 include a local area network (LAN) 971 and a wide  
20 area network (WAN) 973, but may also include other networks. Such networking  
21 environments are commonplace in offices, enterprise-wide computer networks,  
22 intranets, and the Internet.

23 When used in a LAN networking environment, the computer 900 is connected to  
24 the LAN 971 through a network interface or adapter 970. When used in a WAN  
25 networking environment, the computer 900 typically includes a modem 972 or

1 other means for establishing communications over the WAN 973, such as the  
2 Internet. The modem 972, which may be internal or external, may be connected to  
3 the system bus 921 via the user input interface 960, or other appropriate  
4 mechanism. In a networked environment, program modules depicted relative to  
5 the computer 900, or portions thereof, may be stored in the remote memory  
6 storage device. By way of example, and not limitation, Fig. 9 illustrates remote  
7 application programs 985 as residing on memory device 981. It will be  
8 appreciated that the network connections shown are exemplary and other means of  
9 establishing a communications link between the computers may be used.

## Conclusion

The systems and methods as described, thus provide a way to manage interactions from multiple speech-enabled applications, even if two or more of the multiple speech-enabled applications use different grammars. Implementation of the systems and methods described herein provide orderly processing of interactions from multiple speech-enabled applications so a user can more easily communicate with the applications.

Although details of specific implementations and embodiments are described above, such details are intended to satisfy statutory disclosure obligations rather than to limit the scope of the following claims. Thus, the invention as defined by the claims is not limited to the specific features described above. Rather, the invention is claimed in any of its forms or modifications that fall within the proper scope of the appended claims, appropriately interpreted in accordance with the doctrine of equivalents.